

基于深度学习的草图分割算法综述*

王佳欣^{1,2,3}, 朱志亮^{2,3,4}, 邓小明^{2,3}, 马翠霞^{1,2,3}, 王宏安^{1,2,3}



¹(中国科学院大学 计算机科学与技术学院, 北京 100049)

²(计算机科学国家重点实验室(中国科学院 软件研究所), 北京 100190)

³(人机交互北京市重点实验室(中国科学院 软件研究所), 北京 100190)

⁴(华东交通大学 软件学院, 江西 南昌 330013)

通信作者: 马翠霞, E-mail: cuixia@iscas.ac.cn

摘要: 草图一直是人类传递信息的重要工具之一. 草图可以通过简单明了的形式更快地表达人类的一些复杂思想, 因此, 草图处理算法一直是计算机视觉领域的研究热点之一. 目前, 对草图的研究主要集中在识别、检索和补全等方面. 随着研究者对于草图细粒度操作的重视, 对草图分割方面的研究也得到越来越多的关注. 近年来, 随着深度学习与计算机视觉技术的发展, 出现了大量基于深度学习的草图分割方法, 草图分割的精确度和效率也都得到了较大提升. 但是, 由于草图自身的抽象性、稀疏性和多样性, 草图分割仍然是一个非常具有挑战性的课题. 对基于深度学习的草图分割算法进行整理、分类、分析和总结, 首先阐述了3种基本的草图表示方法与常用的草图分割数据集, 再按草图分割算法的预测结果分别介绍了草图语义分割、草图感知聚类与草图解析算法, 然后在主要的数据集上收集与整理草图分割算法的评测结果并对结果进行分析, 最后总结了草图分割相关的应用并探讨未来可能的发展方向.

关键词: 草图分割; 感知聚类; 语义分割; 草图解析

中图分类号: TP391

中文引用格式: 王佳欣, 朱志亮, 邓小明, 马翠霞, 王宏安. 基于深度学习的草图分割算法综述. 软件学报, 2022, 33(7): 2729–2752. <http://www.jos.org.cn/1000-9825/6299.htm>

英文引用格式: Wang JX, Zhu ZL, Deng XM, Ma CX, Wang HA. Survey on Sketch Segmentation Algorithm Based on Deep Learning. Ruan Jian Xue Bao/Journal of Software, 2022, 33(7): 2729–2752 (in Chinese). <http://www.jos.org.cn/1000-9825/6299.htm>

Survey on Sketch Segmentation Algorithm Based on Deep Learning

WANG Jia-Xin^{1,2,3}, ZHU Zhi-Liang^{2,3,4}, DENG Xiao-Ming^{2,3}, MA Cui-Xia^{1,2,3}, WANG Hong-An^{1,2,3}

¹(School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing 100049, China)

²(State Key Laboratory of Computer Science (Institute of Software, Chinese Academy of Sciences), Beijing 100190, China)

³(Beijing Key Laboratory of Human-Computer Interaction (Institute of Software, Chinese Academy of Sciences), Beijing 100190, China)

⁴(School of Software, East China Jiaotong University, Nanchang 330013, China)

Abstract: Sketches have always been one of the important tools for human communication. As it can express some complex human thoughts quickly in a succinct form, the sketch processing algorithm is one of the research hotspots in the field of computer vision. Currently, the research on sketches mainly focuses on the recognition, retrieval, and completion. As researchers focus on the fine-grained operation of sketches, research on sketch segmentation has also received more and more attention. In recent years, with the development of deep learning and computer vision technology, a large number of sketch segmentation methods based on deep learning have been proposed. Moreover, the accuracy and efficiency of sketch segmentation have also been significantly increased. Nevertheless, sketch segmentation is still a very challenging topic because of the abstraction, sparsity, and diversity of sketches. This study organizes,

* 基金项目: 国家自然科学基金(61872346); 国家重点研发计划(2016YFB1001200)

收稿时间: 2020-08-07; 修改时间: 2020-09-02, 2020-10-13; 采用时间: 2020-11-23; jos 在线出版时间: 2021-01-15

categorizes, analyzes, and summarizes the sketch segmentation algorithm based on deep learning to solve the above deficiency. Firstly, three basic sketch representation methods and commonly used sketch segmentation datasets are shown. According to the sketch segmentation algorithm prediction results, sketch semantic segmentation, sketch perceptual grouping, and sketch parsing are introduced respectively. Moreover, the evaluation results of sketch segmentation are collected and analyzed on the primary data sets. Finally, the application of sketch segmentation is summarized and the possible future development direction is discussed.

Key words: sketch segmentation; perceptual grouping; semantic segmentation; sketch parsing

草图是一种原始的数据符号,反映着人的大脑对于世界的认知.草图由人创作,基于人的认知对事物与人的相关信息进行表达,没有任何草图学习经历的人也可以通过触控与笔设备使用草图.图像与文字是传递信息的主要载体,而草图是一种简洁的图像,同时,按文字的图画起源说^[1],草图的出现要早于文字.与其他复杂的图像相比,草图以其固有的简洁性,在表达复杂构思与新型想法时,可以通过更小的数据量传达相近的信息量.与各式各样的语言相比,草图可以基于相同的形式快速表达复杂的思想,避免由于使用不同语言而产生的交流障碍.随着人机交互技术的发展,电子设备与触摸屏得到大力的推广和普及,基于笔设备与触摸屏的人机交互应用也越来越多,手绘草图在计算机上的应用也随之越来越多.目前,手绘草图已经成了人们快速表达如流程图、思维脑图、原型设计等复杂思想的一种常用工具,在汽车设计、工业设计、建筑设计、机械工程设计上发挥着巨大作用^[2].与基于 WIMP (Window, Icon, Menu, Pointer)的交互模型相比,基于草图的交互模型 SBIM (sketch-based interactive model)在表达即时想法时具有更高的效率,可以提高会议与交流的效率,也可以加快 3D 模型设计.依据用途的不同,可以将草图粗略地分为卡通草图与设计草图.卡通草图指主要面向漫画与动画领域的艺术作品,通过简洁、幽默、风趣、夸张、比喻等手法设计的绘画作品;设计草图指主要面向工业与建筑领域的专业作品,具有严格的规范与专业的设计流程.另外,依据维度的不同,可以将草图分为 2D 草图和 3D 草图;依据颜色的不同,可以将草图分为黑白、灰度与彩色草图.

本文主要针对的草图为基于线条的黑白手绘草图(free-hand sketch),即通过人工制作仅由黑白两种颜色的笔画与相应空白构成的草图.图 1 展示了黑白手绘草图与相关图像的区别.第 1 行黑白手绘草图、符号标志图像、边缘图像与图表草图强调线条与形状,其中,黑白手绘草图由存在随机抖动的线条构成,是人脑对于某个物体或概念呈现,如日月星辰等;符号标志图像以颜色、几何形状为主要特征,用以表达某种事物或概念,具有很强的规范性,如常见的天气、交通标志符号;边缘图像呈现物体的轮廓,可以由自然图像通过边缘抽取技术获得;图表草图通过手工绘制的流程图、树状图、统计图等.第 2 行彩色草图、线条艺术图像、卡通图像与自然图像强调颜色与区域,其中,彩色草图图像以颜色、线条与区域为主要特征,常见于水彩画、儿童画等美术作品;线条艺术图像以线条与区域为主要特征,常见于铅笔画、中国画;卡通图像以文字与拟人化、夸张化的事物图像为主要特征,常见于漫画、动画与吉祥物之中;自然图像则是通过相机数字化的真实世界图像,由像素点构成,以颜色、位置与亮度为主要特征.第 3 行草图图案、工业草图、建筑草图与地图草图强调多目标与场景,其中,草图图案图像以点、线、面为主要特征,对少数目标运用对称与重复等手段构成场景,体现出相应的规律性与艺术性以符合人类的审美观;工业草图由多个零件的草图与相应文字规格注释构成,具有严格的制作规范,需要用直尺等工具来避免线条抖动;建筑草图则是由多个建筑物单元与相应的文字规格注释构成,同样需要工具来保证制作的规范性;草图地图是草图与地理地图的结合,带有丰富的地理元素如地形地貌等.除列出的图像外,卫星图像、红外图像与文字图像等也可以与手绘草图相关联.基于黑白手绘草图,经过规范化、组合与注释可以获得符号标志图像、草图地图、工业草图与建筑草图,通过平移扩充、线条上色、区域上色、区域填充与文字添加等步骤,可以获得图案图像、彩色草图、线条艺术和卡通图像等复杂图像,而利用训练好的 GAN (generative adversarial network)可以通过黑白手绘草图或边缘图像生成与编辑自然图像^[3-5].黑白手绘草图与其他图像具有一定的结构与语义相关性,但也有其独有的多样化、稀疏与简洁等特点,因此,针对黑白手绘草图的应用需要依据草图特点进行展开.在本文中,如无特殊说明,“草图”均指“黑白手绘草图”.

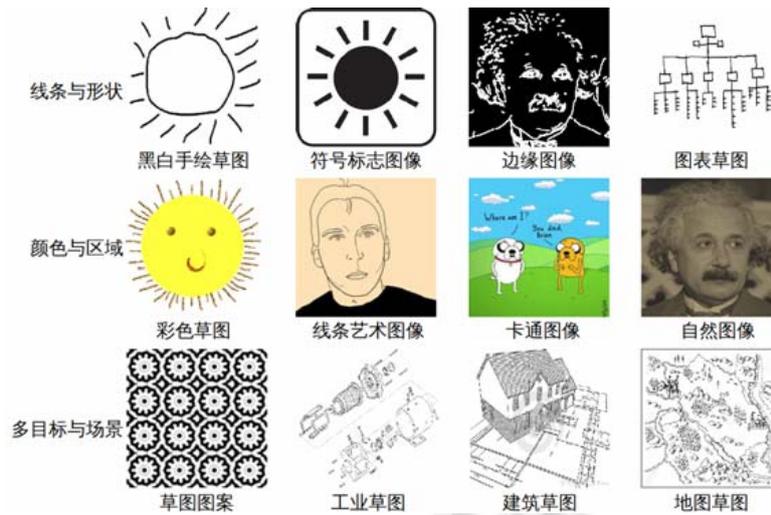


图1 黑白手绘草图与相关图像

与自然图像处理任务类似,草图处理任务也包括分类、识别和分割.与草图分类与识别相比,草图分割支持从笔画层次对草图进行理解,利于理解与编辑用户的意图.与自然图像相比,首先,草图缺乏区域的填充信息,主要包含边缘与轮廓信息,因此,基于区域的图像分割方法并不完全适用于草图的分割;其次,与自然物体的真实轮廓不同,草图注重于基于人类认知的局部主体特征的抽象表示,常常会在某些部位省略诸多细节;最后,由于人类认知水平以及创作风格的不同,比如基于专业性、年龄与性别等差异,同一个物体或概念往往会对应不同的草图.

随着计算机视觉技术的发展,图像的处理由人工设计特征逐渐转为机器自动学习.一方面,由于草图的风格与种类的多样性;另一方面,由于传统的基于人工设计特征的方式难以定义多样的特征,因此,目前的主流是采用深度学习对草图高度抽象的特征进行学习.基于深度学习的草图分割技术可以按处理或添加的主要网络结构分为基于卷积神经网络 CNN (convolutional neural network)、全卷积网络 FCN (full convolutional network)、循环神经网络 RNN (recurrent neural network)和图神经网络 GNN (graph neural network)等.由于 AlexNet^[6], VGG^[7], ResNet^[8]等 CNN 网络在自然图像识别上取得的巨大突破, CNN 网络在 2015 年被引入草图识别^[9],并在 2019 年应用于草图分割^[10];由于 FCN^[11], UNet^[12], DeepLab 系列^[13-16]等在自然图像语义分割上取得突破, FCN, UNet 与 DeepLab 等被广泛用作草图分割的基础实验参考对象(baseline); RNN^[17]与 LSTM (long short-term memory)^[18]在自然图像目标识别^[19]、自然图像/视频自动标题(image/video caption)^[20,21]上的应用,启发了 2016 年将 RNN 应用于草图识别^[22]与 2018 年将 RNN 应用于草图补全^[23]的工作;生成对抗网络 GAN^[24]在自然图像风格转换(image-to-image transfer)、自然图像超分辨率(image super resolution)、自然图像补全(image completion)和从文本生成自然图像(text-to-image generation)上的应用^[25],启发了 2016 年将 GAN 应用于草图检索^[26]与 2018 年将 GAN 应用于从草图到图像的合成(sketch-to-image synthesis)^[4]的工作; GNN^[27]在场景图生成、点集分类和动作识别上的应用^[28],启发了 2019 年将 GNN 应用于草图识别^[29]与 2020 年将 GNN 应用于草图笔画语义分割^[30]的工作.

基于深度学习的草图分割技术为草图应用提供更高精确度的细粒度图像理解支持,使得更符合用户理解的草图编辑、交互、补全与推荐等成为可能.如图 2 所示,草图分割依据分割的最小单元,可以分为笔画级标注与像素级标注.笔画级标注对构成草图的每一笔画进行分类,专注于构成草图的笔画,对绘制人员的每一笔笔画进行分类或分组,因而会忽略草图中的空白;另外,对于语义区域交界处存在一定歧义性的笔画,如果仅赋予一个标签或者仅分到某一组是不合理的.像素级标注对草图图像中的每一像素进行标注,不仅对构成草图的笔画进行标注,也对草图图像中的空白区域赋予标注,与真实的场景更加接近,它将笔画视为语义

区域之间的分界线,可以避免笔画级标注的歧义性.场景语义分割将场景草图划分成不同的目标,而单目标语义分割将单目标草图划分成不同的组件,如将马分为头、角、身体与四肢这 4 个组件并赋予相应的语义标签,而单目标感知聚类将目标草图的笔画进行分组,但不赋予相应的语义标签.草图解析则对目标草图中的所有像素赋予标签,将图像划分为不同的语义区域.手绘草图分割技术依据语义或感知理论对草图进行划分,可以广泛应用关注局部细节、区域或组件的应用,如草图编辑与交互、基于草图的细粒度图像生成、检索与编辑、基于草图的 3D 模型合成与检索等应用,但是目前国内缺乏关于手绘草图分割的综述.针对这个不足,本文将对近年基于深度学习的手绘草图分割相关研究工作进行综述.

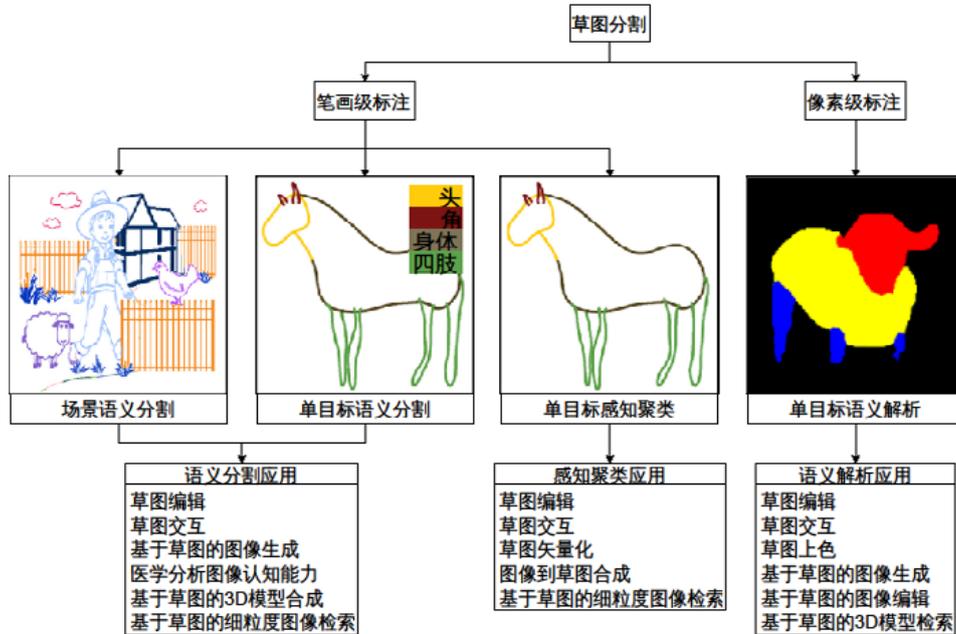


图 2 草图分割的分类与应用

1 草图的表示

1.1 基于图像表示方法

草图作为自然图像的一种表现形式,可以采用常用图像的方式进行表示^[9].一幅草图 I 可以表示为 2 位的黑白图, 8 位的灰度图或者 24 位的 RGB 彩色图. 即

$$I = A_{mn}^d, d \in \{2, 8, 24\}, \quad (1)$$

其中, m 与 n 分别为图像的高度与宽度, A 为与图像等高等宽的矩阵, d 为每一个矩阵元素所占的比特位数. 显然, 本文针对的基于黑白线条的手绘草图最简洁的图像表示方法为黑白图, 但为了利用从自然图像上预训练的特征, 可以采用相对复杂的 RGB 彩色图表示. 基于图像表示方法可以使得每个草图占用相同的内存, 便于应用自然图像处理算法, 但会丢失绘制草图的时序信息, 并因空白区域而浪费内存.

1.2 基于序列表示方法

根据 TU-Berlin 数据集的统计规律, 人们在绘制相同物体的草图时, 倾向于按照相同的顺序, 如在画一张人脸时, 人们倾向于先画出人脸的轮廓, 再依次画左右眼与鼻子嘴巴^[31]. 而基于图像表示方法丢失了草图的这种时序信息, 不利于对草图的处理. 草图在电子设备上进行输入时, 通常可以通过窗口坐标得到当前笔头在画布上位置. 在这种可以获得时序信息的情况下, 草图适合以序列的方式进行表示, 如 $I = [t_1, t_2, \dots, t_i, \dots]$, 其中, 某一时刻 i 的位置与状态通过变量 $t_i = [p_i, s_i]$ 表示, 当前位置信息通过坐标 $p_i = [x_i, y_i]$ 表示, 当前笔画(stroke)

状态通过变量 s_i 表示. 草图生成(sketch generation)模型 Sketch-RNN^[23]扩展了文献[32]中草图的序列表示方法, 通过位置的偏移量 $p_i=[\Delta x_i, \Delta y_i]$ 和画笔的状态 $s_i=[\phi_i^0, \phi_i^1, \phi_i^2]$ 来表示草图的笔画. 其中, ϕ_i^0 表示画笔是否接触到画纸, ϕ_i^1 表示画笔是否会在当前位置离开画纸, ϕ_i^2 表示绘画是否结束. 该序列表示方法便于应用循环神经网络 RNN, 因此被广泛应用于其他草图应用如草图合成(sketch synthesis)^[33]、草图聚类(sketch grouping)^[34]和草图摘要(sketch abstraction)^[35].

$$s_i = \begin{cases} [1, 0, 0], & \text{画笔接触纸} \\ [0, 1, 0], & \text{画笔离开纸} \\ [0, 0, 1], & \text{绘制结束} \end{cases} \quad (2)$$

1.3 基于点集的表达方法

基于序列的方法具有不确定的长度, 通常需要通过填补标志信息使得序列等长, 从而应用基于序列的算法. 这种人为引入冗余的方式不利于算法进行学习处理, 因而, 文献[30]通过点集来表示整个草图, 即 $I=\{p_1, p_2, \dots, p_i, \dots, p_N\}$, 其中, 位置的绝对坐标 $p_i=[x_i, y_i]$; N 为草图依据算法^[36]的采样点数, 在原文中设置为128或256. 基于点集的表达方法忽略笔画内与笔画间的时序信息, 将整个草图当成点集进行处理.

1.4 草图表示方法的扩展

一方面, 可以将草图的基本表示方法进行组合, 充分利用每种表示方式的优势, 如 SPFusionNet^[37]同时采用图像与点集的表达方式, 融合像素特征与点集特征对草图进行分割; 另一方面, 由于硬件设备的不同, 一些电子设备还可以得到一些额外的草图信息, 如笔头处的压力与朝向, 这种情况可以在草图的图像、序列和点集表示方法上进行扩展, 如增加图像表示笔头信息或者扩展位置信息为 $p_i=[x_i, y_i, u_i, v_i, \dots]$, 从而表示额外的草图信息.

2 草图分割相关数据集

2.1 TU-Berlin

TU-Berlin^[31]数据集公开于2012年, 依据给定的日常物体类别如猫、茶壶, 收集来自1350个非专家绘制出的草图. 它包含250个独立的物体类别, 几乎包含了所有的日常物体, 平均每一幅草图含有13条笔画, 平均每一类含有80张草图, 总计20000张手绘草图. 其中的草图仅通过形状信息即可得到其类别, 不需要上下文信息. Schneider等人^[38]从TU-Berlin中随机选择了6类, 每类手动标注了20个样本, 构成了Schneider数据集, 总计120张草图用于实验验证. Im2Sketch^[39]从TU-Berlin中选择了96类, 通过15个标注者进行手动标注, 每类80个样本, 构成了TU-Berlin-Qi数据集, 总计7680张草图. EdgePG^[40]从TU-Berlin中选择了25类, 构成感知聚类数据集TU-Berlin-EdegPG. 如图3所示, TU-Berlin数据集中的草图所含笔画较少, 部分草图如手臂(arm)与苹果(apple), 所含笔画可能少于5笔; 而Schneider数据集从TU-Berlin中随机选取的6类, 笔画相对较多.

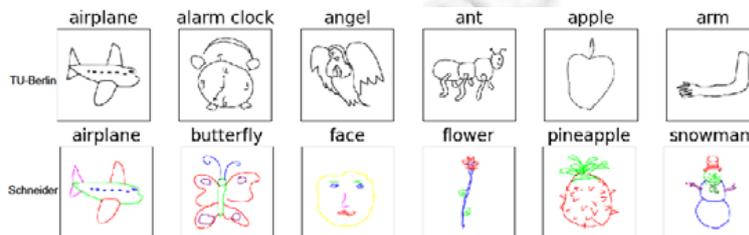


图3 TU-Berlin与Schneider数据集示例, 其中, 不同的颜色表示不同的标签

2.2 Huang

Huang^[41]数据集公开于2014年, 包含10类草图, 对于每一类, 由一个专家用户和两个普通用户绘制收集.

800 张, 总共 20 000 张草图. 所有的草图都经过草图识别算法^[9]过滤, 无法识别的草图将不会被选择. 同时, 基于复杂性(草图超过 3 个语义部件)、多样性(仅有一种四足动物被选择)、唯一性(某个大类包含多种子类, 则只选择其主要的子类草图)和模糊性(草图不能包含其他物体或者背景)对草图进行筛选. 草图由 25 个人基于少数服从多数的原则进行标注. 如图 6 所示, SPG 数据集中的草图容易识别对应的类别与语义.



图 6 SPG 数据集示例, 其中, 不同的颜色表示不同的标签^[43]

2.5 SketchSeg

SketchSeg-57K^[44]数据集公开于 2018 年, 基于 QuickDraw 数据集, 通过从中选择 7 类, 每类随机选择 60 张以上的草图, 并通过人工标注其语义部件, 再利用 Sketch-RNN 生成其余草图, 最终获得 57 000 张草图与语义部件标注. 如图 7 所示, SketchSeg-57K 数据集中包含不同语义标签数的草图, 同一物体的草图随着笔画数的增加, 其语义标签也可能增加, 如咖啡杯(coffee cup)从两类变到 4 类. SketchSeg-150K^[45]数据集公开于 2019 年, 基于 QuickDraw 数据集, 包含 20 类. 通过人工标注少量选择的草图, 再通过 Sketch-RNN 生成其余草图, 得到共 150 000 张草图与相应的语义部件标注. SketchSeg-10K^[37,46]数据集公开于 2019 年, 含有 10 类, 每类 1 000 张草图, 共计 10 000 张草图与对应的像素级标注, 由 4 个专业人员与 16 个非专业人员在不提供参考图像的情况下绘制而成, 每个人为每一类绘制 50 张草图.

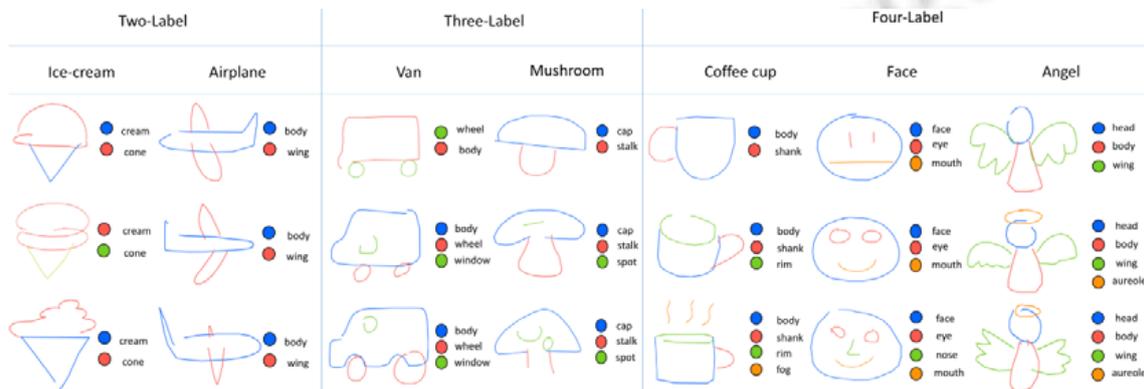


图 7 SketchSeg-57K 数据集示例, 其中, 不同颜色表示不同的标签^[44]

2.6 SemanticPart

SemanticPart^[47]数据集公开于 2019 年, 针对草图语义的特点进行研究. 它由 326 名参与者标注, 每名参与者从一个已有草图识别数据集^[48]中随机选择 10 张进行笔画标注. 由于参与者可以创造新的标签, 由于一共产

生了 304 个不同的标签, 经过预处理, 将一些含义相近的标签映射到相同的标签上, 如将椅子的标签脚(leg)、足(foot)、支柱(strut)、支撑腿(leg support)等映射到相同的标签腿(leg)上, 得到 24 个映射后的标签. 经过预处理, 最终数据集包含 864 张图像, 分为 4 类(鸟、车、椅、狗), 并被标注了 3 次. 对数据集的研究发现: (1) 人们对单个笔画的理解是一致的, 95.6% 的线条得到了至少 2 人以上的相同标签, 67.8% 的线条得到了 3 个参与者相同的标注; (2) 单个笔画往往对应单个标签, 同时, 相同标签的笔画绘制的时间也相近; (3) 同一物体的复杂草图与简单草图强调相似的部件信息. 如图 8 所示, 第 1 排为草图对应的自然图像, 第 2 排为复杂草图, 第 3 排为简单草图.

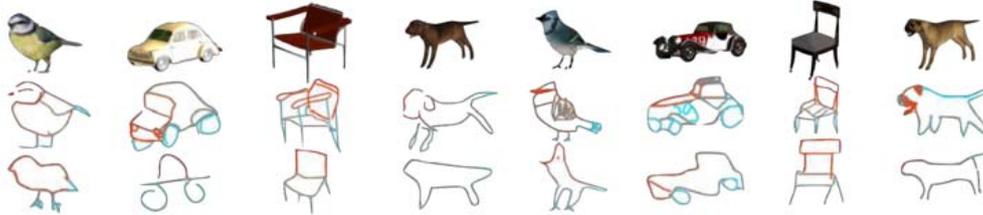


图 8 SemanticPart 数据集示例, 其中, 不同的颜色表示不同的标签^[47]

2.7 SketchParse

SketchParse^[49]数据集公开于 2017 年, 为草图解析而构建, 通过自然图像生成草图进行训练, 并在草图上进行验证. 其训练集具有 11 种类别, 共包含 1 532 对自然图像与通过数据增广生成的相应 21 448 张草图及组件标注, 自然图像从 Pascal-Part^[50]与 Core^[51]数据集中选取; 其测试集包含 11 类, 共 1 008 张草图, 与自然图像的 11 类对应, 草图从 Sketchy^[52]与 TU-Berlin 数据集中选取, 如图 9 所示, SketchParse 数据集标注侧重区域, 存在标注的噪声, 与草图相比, 标注结果忽略部分草图笔画. 如自行车(bicycle)的标注忽略了部分自行车支架, 而猫(cat)的标注忽略了面部的线条.

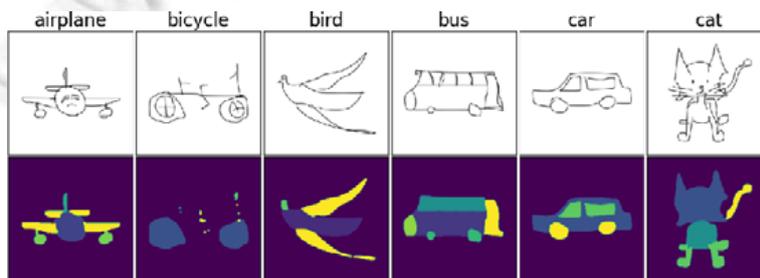


图 9 SketchParse 数据集示例, 其中, 不同的颜色表示不同的标签

2.8 SketchyScene

SketchyScene^[53]数据集公开于 2018 年, 是第一个大型的场景草图数据集, 通过交互程序构建, 基于给定模板, 用户通过选择与绘画补充单个目标草图, 最终合成场景草图. 它包含 29 000 张以上的场景草图, 7 000 对以上的场景模板与图像, 11 000 张以上的单个目标草图, 所有场景草图均具有语义与实例级标注. 如图 10 所示, 数据集中的草图场景包含常见的物体类别, 存在相互遮挡与尺度变化等挑战.



图 10 SketchyScene 数据集示例, 其中, 不同的颜色表示不同的类别

2.9 小 结

随着草图分割应用需求的不断增加, 涌现出越来越多的草图分割数据集. 表 1 列出了常用的草图分割数据集与相关的草图数据集, 其中, TU-Berlin 数据集是最早的公开草图数据集, 常被二次标注加工成为验证草图分割应用的数据集, 如 Schneider 数据集, 标注 TU-Berlin 数据集中的部分草图样本. Huang 数据集是出现最早的公开草图分割数据集, 大多数草图分割算法都会在较早出现的 Huang 数据集和 Schneider 数据集上进行验证, 尽管这两个数据集对于基于深度学习的模型而言数据量都不足. QuickDraw 数据集从数据集图片量或种类数量上而言是目前已知的最大草图数据集, 可以用来预训练草图分割模型中的特征提取模块; 同时, 它也被 SPG 数据集、SketchSeg-57K 数据集与 SketchSeg-150K 数据集用于二次标注加工. SPG 数据集中所有的草图分割标注都由人工获得; 而 SketchSeg-57K 与 SketchSeg-150K 采用人工获得部分草图分割标注, 再由 Sketch-RNN 模型生成其他草图分割标注. SemanticPart 数据集是由用户自定义标签的草图分割数据集, 可以用来研究不同用户的标注倾向, 虽然仅有 4 类, 但由 325 名人员标注. SketchParse 数据集是当前唯一的公开草图解析数据集, SketchyScene 数据集是当前唯一的公开场景语义分割数据集. 由于草图语义分割数据集同时包含语义标签与分组信息, 因此同样可以用于感知聚类算法的评测.

表 1 草图分割数据集

数据集名称	数据集图片量	种类数量	标注类型	公开时间(年)
TU-Berlin ^[31]	20 000	250	草图识别	2012
Huang ^[41]	300	10	语义分割	2014
Schneider ^[38]	120	6	语义分割	2016
SketchParse ^[49]	21 448	11	草图解析	2017
QuickDraw ^[23]	50 000 000	345	草图识别	2018
SketchyScene ^[53]	29 000	45	场景语义分割	2018
SketchSeg-57K ^[44]	57 000	7	语义分割	2018
SketchSeg-10K ^[37,46]	10 000	10	语义分割	2019
SketchSeg-150K ^[45]	150 000	20	语义分割	2019
SPG ^[43]	20 000	25	语义分割	2019
SemanticPart ^[47]	864	4	语义分割	2019

3 基于深度学习的草图分割算法

3.1 草图分割算法分类

如图 2 所示, 草图分割算法按照输出结果可以分为语义分割、感知聚类与草图解析. 其中, 语义分割为草图中的笔画分配某个拥有具体语义的标签, 按照分割粒度与语义标签的不同, 可以将语义分割进一步分为场景级语义分割与单个目标语义分割; 感知聚类对草图中的笔画进行分组, 使得其结果符合人的感知; 草图解析则为草图图像中的所有像素分配语义标签.

由于本文针对基于深度学习的草图分割算法, 因此对于非深度学习方法与针对草图相关图像(如图 1 所示)的分割算法, 仅挑选相关性强的算法进行介绍. 依据图 2 中的分类方法, 可以将已有的草图分割算法整理为表 2. 其中, 算法集中在 2014–2020 年, 草图语义分割算法占大多数, 分为非深度学习方法与基于图像、序列和点集的深度学习方法. 其中, 基于数据驱动的 Huang-Auto 算法与基于条件随机场的 Sketch-CRF 算法由于提出较早, 常作为其他算法的对比算法; 基于图像的深度学习算法结合草图的特性修改深度神经网络结构, 如 Dual-CNN 采用两个不同卷积核的 CNN; 基于序列的深度学习算法通过修改与改进 Sketch-RNN 草图生成模型, 从而对输出草图分割预测结果; 基于点集的深度学习算法利用点集网络与图卷积网络, 学习草图的图结构特征, 对草图进行分割. 草图聚类算法由于需要获得笔画与笔画之间的相似度, 而采用图像与点集的方式预测像素与像素或者点与点之间的相似度存在巨大的计算开销, 因而适合采用基于序列的方式. 草图解析算法需要预测所有像素点的语义标签, 因而适合采用基于图像的方式.

表 2 草图分割算法

语义分割				
算法分类	算法名称	年份	评测数据集	特点
非深度学习方法	Huang-Auto ^[41]	2014	Ref.[41]	数据驱动
	Sketch-CRF ^[38]	2016	Ref.[38,41]	条件随机场
基于图像	Dual-CNN ^[54]	2018	Ref.[38,41]	两个不同卷积核的 CNN
	FastSeg ^[55]	2019	Ref.[38,41]	CNN 分割后采样优化
	2DSeg ^[10]	2019	Ref.[38,41]	输入笔画与草图图像
	SFSegNet ^[56]	2019	Ref.[37,41]	空间变换模块处理形变
	Sketch-VGG ^[57]	2020	Ref.[38,41]	输入笔画与草图图像, 迁移学习
基于序列	SketchSegNet ^[44]	2018	Ref.[38,41]	基于 Sketch-RNN 与全连接分类器
	SketchSegNet+ ^[45]	2019	Ref.[38,41]	基于 Sketch-RNN 与全连接分类器
	USPG2.0-Pretrain ^[43]	2019	Ref.[38,41,43]	基于 Sketch-RNN 与感知聚类
	Kaiyrbekov-Seg ^[42]	2020	Ref.[41,42]	基于 Stroke-RNN 进行重建与分割
基于点集	MCPNet ^[46]	2018	Ref.[41]	多个不同卷积核的点集网络融合
	Sketch-GCN ^[30]	2020	Ref.[38,41,43,44]	图卷积网络
基于点集与图像	SPFusionNet ^[37]	2019	Ref.[37]	融合图像与点集特征
草图聚类				
算法分类	算法名称	年份	评测数据集	特点
非深度学习方法	Edge-PG ^[40]	2015	Ref.[40]	基于 Rank-SVM 与图模型
基于序列	USPG1.0 ^[34]	2018	Ref.[34]	局部与全局聚类损失
	USPG2.0 ^[43]	2019	Ref.[34,40]	局部与全局聚类损失
草图解析				
算法分类	算法名称	年份	评测数据集	特点
基于图像	SketchParse ^[49]	2017	Ref.[49]	通用、专家与路由网络
	DeepSSP ^[58]	2019	Ref.[49]	齐次变换与迁移学习

3.2 笔画级标注算法——草图语义分割

草图笔画级标注算法将草图中的笔画分类成预先定义的几种语义, 即仅对草图中的笔画进行分类, 忽略草图中的空白区域. 对比自然图像的语义分割算法, 草图的基本元素为笔画, 对草图的语义分割可以对应到对笔画的语义分割, 因此, 基于笔画的草图标注算法在许多文章中称为草图语义分割. 由于不同文章对草图分割算法的命名不同, 本文采用综述文献[59]中的命名, 将草图笔画的语义分类称为草图语义分割, 将草图笔画与空白的语义分类称为草图解析, 将草图笔画的分组称为草图聚类. 基于深度学习的草图语义分割算法, 从表示方式上大致可分为基于图像、基于序列和基于点集.

3.2.1 早期草图语义分割相关方法

早期方法采用人工设计的特征对草图或类似图像如边图与剪贴画等进行分割, 如文献[60–62]将草图分割为基本的几何结构如直线、圆和圆弧, 其中, Sezgin 等人^[60]利用笔画的方向、速度与曲率特征, 预测分割点从而分割草图; Kim 等人^[61]利用曲率进行分割; Pu 等人^[62]利用径向基函数(radial basis function), 结合方向与曲率对草图进行分割. Sun 等人^[63]提出了首个从场景上对草图进行分割的方法, 基于自定义的特征, 计算笔画之间的距离, 从而将剪贴画分割为不同的物体. Noris 等人^[64]基于位置、时序与交互时引入的草图标注信息, 建立了 MRF 模型进行草图分割. Huang 等人^[41]通过数据驱动的方式解决草图分割问题, 输入一张草图, 首先利用卷积神经网络学习得到匹配点, 再通过聚类和点集匹配对相应点进行预处理, 最后将草图聚类问题转换为混合整数规划问题(mixed integer programming problem)进行求解. 该方法由于存在多个阶段, 难以进行全局优化; 同时, 匹配点的计算与预处理方式难以保证准确性. Schneider 等人^[38]利用人工设计的 SIFT 特征^[65]与 Fisher Vector^[66]特征构建图, 通过图划分将笔画在曲率大的位置进行分割, 并通过条件随机场(conditional random field)^[67]提取草图的图结构信息, 优化分割结果.

3.2.2 基于图像的草图语义分割算法

基于图像的草图笔画语义分割算法将草图视为自然图像, 将针对自然图像语义分割的算法应用到草图上^[14]. 文献[10,54,55,57]依据草图的特点对自然图像的语义分割网络进行改进, 但此类算法缺乏笔画之间的

结构信息, 在输出最终分割结果前, 需要通过建立条件随机场 CRF 等方式引入结构信息的图模型, 并通过图割算法^[68,69]进行求解优化. Zhu 等人^[10]通过定制卷积神经网络, 使第 1 个卷积层具有更大的卷积核与卷积步长 (stride), 从而更好地捕捉草图稀疏的笔画特征. 同时, Zhu 等人^[10]提出了 HostSketch 模型, 分别将各条笔画与整张草图的图像作为输入, 综合局部笔画信息与全局类别信息, 基于 Sketch-a-Net 草图分类模型^[9], 单独对每笔笔画进行分类. 与此类似, Zhu 等人^[57]基于自然图像分类网络 VGG16 进行修改与精调, 提出了模型 Sketch-VGG, 将草图以某个笔画、整张草图、空白图像的形式输入网络, 从而保持网络的输入通道数不变. HostSketch 与 Sketch-VGG 模型的运行时间依赖于草图笔画的数目, 同时需要草图的笔画分割信息来预测每条笔画的语义类别.

Li 等人^[55]提出了基于编码解码器(encoder-decoder)的 FastSeg 模型, 通过已有 3D 模型数据库生成草图数据集, 并训练神经网络对草图进行语义分割, 解决了手绘草图语义分割缺少数据的问题. 但由 3D 模型生成的数据集在多样性上不及人工绘制的草图, 缺少绘制者由个人的风格与技术而带来的差异, 与手绘草图存在明显的差别. 因此, 基于 3D 模型生成的草图训练模型推广到手绘草图应用领域的结果难以预料. 该文在神经网络分割结果的基础上, 通过点采样与图割方法^[68]优化, 在 Huang 数据集上, 基于部件的精度(component-based accuracy)由 65.6%提升到 73.8%, 提高了 8.3%; 在 TU-Berlin 数据集上, 基于部件的精度由 71.0%提升到 79.4%, 提高了 8.4%. 然而, 由于训练数据与真实手绘草图间的差异, 该文方法会因视角或异于生成数据集中的笔画特点而导致分割错误. Jiang 等人^[56]提出了一个针对草图的全卷积分割网络 SFSegNet, 如图 11 所示, 该网络由修改后的全卷积网络 FCN 与仿射变换编码器构成, 采用重加权的策略忽略背景像素, 从而避免前景背景样本不均匀的问题, 并通过仿射变换规范笔画抖动, 从而解决扭曲笔画的问题. 由于输入的草图以 RGB 图像表示, 因此该文部分网络结构在 ImageNet 图像数据集上进行了预训练, 并在相应草图数据集上进行精调 (fine-tune). 仿射变换编码器由卷积层、池化层与两个全连接层构成, 生成带有 6 个参数的仿射变换矩阵, 对来自编码器中的特征进行变换, 达到空间变换网络(spatial transformer network)^[70]的效果, 并在 SketchSeg-10K 数据集上^[37]进行对比实验验证, 其中, *P-metric* 由 89.1%提升到 91.2%, *C-metric* 由 86.5%提升到 88.6%.

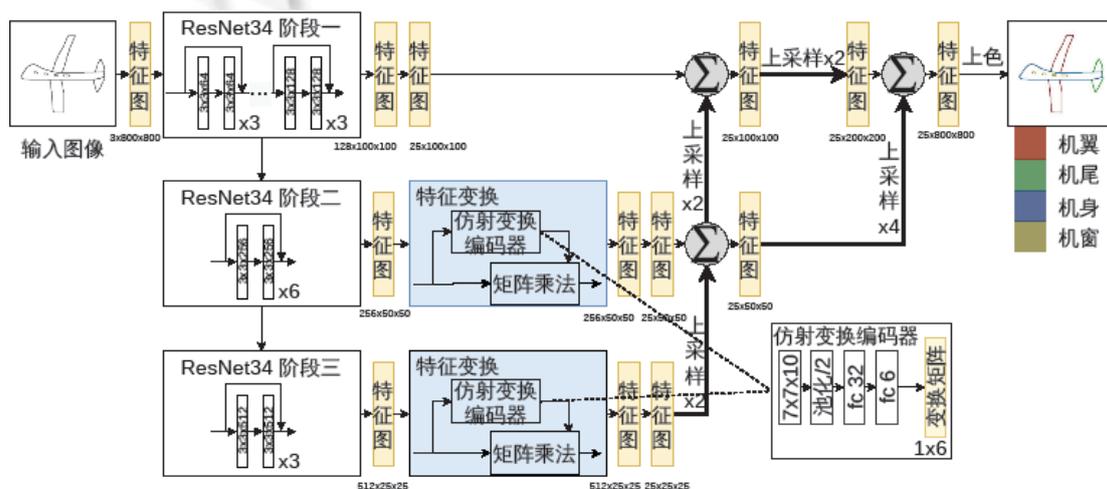


图 11 SFSegNet 网络结构图^[56]

Zou 等人^[53]首次提出了草图场景分割数据集 SketchyScene, 并利用它训练基础模型 DeepLab-v2, 对草图场景进行语义分割, 并将它与 FCN-8s, SegNet^[71], DeepLab-v3 等自然图像语义分割模型进行对比. 实验发现: 由于草图的稀疏性与多样性, DeepLab-v3 所引入的上下文信息对于草图分割的提升有限, DeepLab-v2 更适合应用于草图分割.

3.2.3 基于序列的草图语义分割算法

基于序列的算法将草图分割视为序列预测问题^[43-45], 通过将草图编码为点的位置及笔画的状态, 再通过循环神经网络进行处理. Wu 等人^[44]首次将循环神经网络用于草图笔画分割, 基于 Sketch-RNN 草图生成模型, 将经过双向 LSTM 编码后的特征输出到 LSTM 解码器, 构建基于草图笔画的 SketchSegNet 分割模型. 如图 12 所示, 该方法首先基于双向循环神经(bidirectional recurrent neural network)^[72,73]的编码器将草图笔画编码为矢量化序列, 再利用基于变分自动编码器(variational autoencoder)的序列到序列模块(sequence-to-sequence model)^[74,75]来学习笔画的顺序模式, 进一步生成隐变量向量作为解码器的初始隐藏状态; 然后, 基于自回归循环神经网络(autoregressive RNN)^[76]的解码器, 将之前的隐藏状态与待分类的笔画作为输入, 并编码相应笔画特征; 最后, 通过全连接网络与 Softmax 激活层得到该笔画的分类结果. 对按顺序与逆序输入的笔画序列, 通过单向 LSTM^[77]按顺序输出其预测的分类结果. 将每一个笔画的分类结果进行综合, 即可得到对整个草图的语义分割结果. 与 Sketch-RNN 草图生成模型相比, SketchSegNet 草图分割模型将解码器最后的高斯混合模型 GMM (Gaussian mixture model)采样模块更换成由全连接层与 Softmax 激活层构成的分类模块. Qi 等人^[45]在 SketchSegNet 模型的基础上提出了 SketchSegNet+模型, 其中, Softmax 层输出 57 类组件标签的概率, 增强后的模型可以对多类草图进行分割, 给实际应用带来便利. 基于序列的算法在一定程度上缺乏不同笔画在全局上的分布信息, 而缺失这种分布信息将难以正确理解草图; 与此同时, 该算法需要在输入中提供笔画的顺序、起点、终点和长度, 无法处理仅以图像形式呈现的草图.

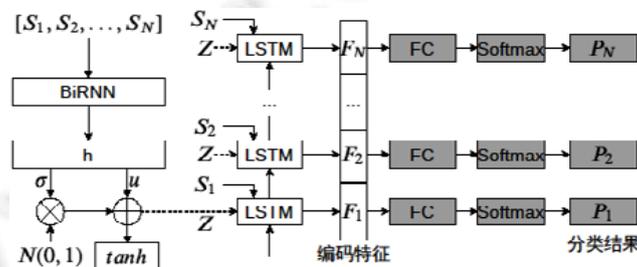


图 12 SketchSegNet 网络结构图^[44]

3.2.4 基于点集的草图语义分割算法

基于点集的算法充分考虑了草图的稀疏性, 利用点集处理网络或图卷积网络对草图进行处理, 但忽略了草图笔画的顺序信息. 基于点集处理网络的算法可以避免对二维卷积核不合理的使用与学习, 从而充分利用草图的稀疏性. 如 MCPNet^[46]受 PointNet^[78]启发, 将输入草图通过采样转换为二维点集, 再通过不同大小的卷积核与多列结构学习与整合草图的多尺度特征表达, 最终生成分割图. 实验证明, 多列结构有助于提高分割结果的精度. 而 SPFFusionNet^[37]综合点集与图像两种表示方式, 一方面采用 SketchNet 学习草图图像的多层次像素特征, 另一方面采用 SPointNet 学习草图点集的全局与局部点集特征, 最后整合像素与点集特征预测分割结果. 其中, SketchNet 采用编码解码结构, 结合空间不变性增强残差(spatial invariance enhanced residual)模块处理草图的形变, 学习并编码草图的多尺度像素特征. 而 SPointNet 同样受 PointNet 启发, 学习并编码草图的全局与局部点集特征.

基于图卷积网络的算法, 需要考虑动态边与静态边. 动态边与静态边相比具有更好的图结构学习能力, 但可能引入错误. 如图 13 所示, SketchGCN^[30]将输入草图以点集进行表示, 通过构建相应的图结构, 利用多层动态卷积模块与静态卷积模块学习草图的动态图特征与静态图特征, 再通过混合池化模块与拼接综合动态边与静态边特征, 在保证学习稳定的同时, 提高图神经网络的学习能力与草图笔画语义的分割性能.

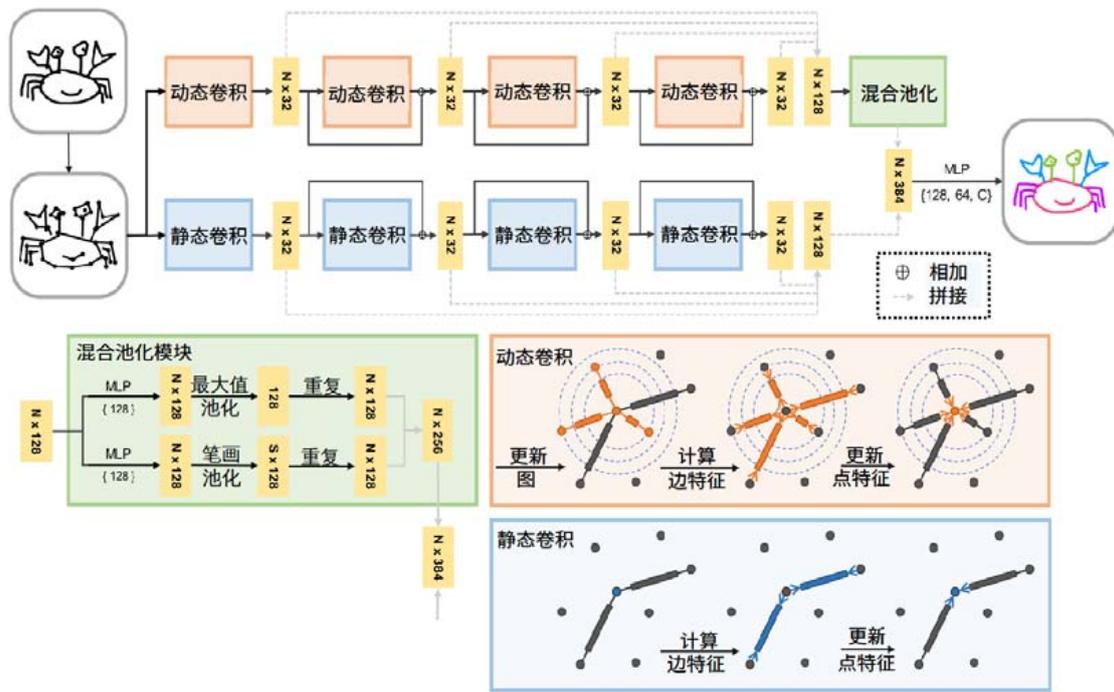


图 13 SketchGCN 网络结构图^[30]

3.3 笔画级聚类算法——草图感知聚类

草图笔画级聚类通过将草图上的笔画或点进行分组,使得每组中的元素具有相近的意义,形成草图的子结构,如将一个人脸草图分组成五官眉眼耳鼻口.文献[79-81]认为,人类识别物体与物体组件的能力,即将视觉线索与图案进行感知分组(perceptual grouping)的能力建立在格式塔分组法则(Gestalt laws of grouping)上,如图 14 所示,包括相邻性、相似性、连续性、闭合性与对称性等.草图聚类的目标是与人类划分的结果一致,因此,草图笔画级聚类也称感知聚类(即感知分组).格式塔分组法则难以用准确的数学语言进行定义,同时,其分组法则之间也可能存在冲突^[82],因此,草图聚类其主要难点有:(1)格式塔分组原则在草图上的表示;(2)对不同分组原则进行组合应用;(3)依据草图的特性学习或设计草图特征,定义笔画之间的距离;(4)在聚类数目未知的前提下,给出符合人感知的聚类结果.基于深度神经网络进行特征学习的草图聚类算法,学习到的原则主要是相邻性与相似性,因而学习到的特征难以保证:(1)通用性,即对新的草图目标进行正确分组;(2)草图笔画对局的部一致性与多个笔画的全局一致性,即如果局部上判定笔画 S_1 与笔画 S_2 为同组,笔画 S_1 与笔画 S_3 为同组,那么在全局上不能出现笔画 S_2 与笔画 S_3 不同组的预测结果.

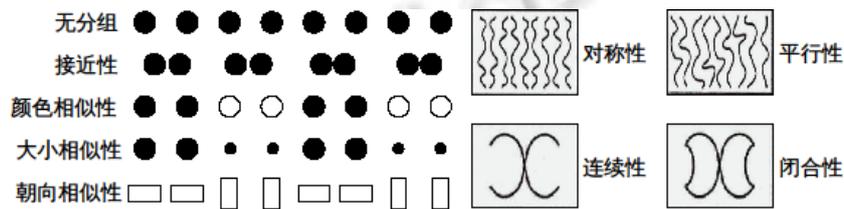


图 14 格式塔感知分组法则^[70]

3.3.1 早期草图聚类相关方法

尽管基于深度学习的草图聚类算法成为目前的主流,但早期的非深度学习草图聚类算法与针对图案、边缘提取结果的聚类算法的基本思想,如对于特定感知分组特征的表达与学习、对于感知分组原则冲突的处理,

可以为之后的深度神经网络模型提供思路. Qi 等人^[83]基于人工设计的格式塔法则特征建立图模型, 并通过图割(graph cut)算法对边缘图或草图进行感知聚类, 并应用到由草图到边图再到合成草图的流程中. Lindlbauer 等人^[84]基于相邻性、端点连续性、平行性、相似性计算笔画之间的距离, 再通过层次聚类算法 HAC (hierarchical agglomerative clustering) 获得草图聚类结果, 并辅助用户对同组草图笔画的选择. 由于草图与自然图像边缘提取结果边缘图的相似性, 用于边缘图感知聚类的算法^[40,85-89]一般也可以应用到草图聚类. Qi 等人^[40]采用 RankSVM 计算感知聚类的相邻性与连续性原则的权重建立图模型, 并通过多标签图割算法^[69]对草图进行分组. Qi 等人^[39]通过 RankSVM 学习格式塔原则的权重来解决格式塔原则之间的冲突, 并基于连续性与相邻性设计人工特征, 建立多标签图割模型并进行求解, 从而获得最终的聚类结果.

3.3.2 针对离散图案的聚类算法

离散图案与草图场景具有许多相似之处, 而据本文调研, 目前还没有针对草图场景进行聚类的算法, 基于离散图案的聚类算法对草图场景聚类具有重要参考价值. Lun 等人^[90]基于感知聚类的原则提出了第一个对离散图案进行感知聚类的深度卷积神经网络 PatternPG. 它由基本元素编码器(atomic element encoder)与结构编码器(structure encoder)构成, 其中, 基本元素编码器对离散图案中的基本图案进行编码, 结合结构编码器提供的全局位置与尺寸信息, 得到最终的融合特征用于离散图案感知聚类. PatternPG 与基于深度学习的草图聚类算法目标类似, 但主要针对由基本图案构成的离散图案. 类似工作如 Greff 等人^[91]则基于期望最大化算法, 利用 TAG 网络对纹理和形状进行迭代感知聚类.

3.3.3 草图矢量化算法

草图矢量化算法对于草图聚类具有扩展与启发作用, 该算法不仅对草图中的可见笔画进行聚类, 还需要预测被遮挡区域的笔画. 比如, Kim 等人^[92]提出全卷积网络 PathNet 找出图像中像素点的所在路径, 并提出与 PathNet 结构类似的全卷积网络 OverlapNet 预测重叠区域, 最后利用 MRF 对所有信息进行综合, 并通过图割算法 GraphCut 进行求解. 该方法不仅可以实现对草图的分割, 还可以恢复草图被遮挡的区域. 另外, 该方法在 3 种不同的数据集(中文和日文字符、合成随机线条和 QuickDraw 手绘草图)上进行了验证, 并取得了平均 IOU 超过 0.8 的良好结果. 但由于 PathNet 需要运行多次来预测笔画之间的相关性, 用于优化的 GraphCut 算法要迭代多次, 总处理时间在一些草图上超过 400 s, 难以达到实时应用的需求.

3.3.4 基于深度学习的草图聚类算法

USPG1.0^[34]收集了大型草图聚类数据集 SPG, 并基于 Sketch-RNN 草图合成模型提出了第一个基于深度神经网络的通用草图聚类模型 USPG1.0. 如图 15 所示, USPG1.0 将 Sketch-RNN 作为其生成网络分支, 通过序列到序列的变分自动编码器(sequence-to-sequence variational autoencoder)^[74,75]学习草图的模式. USPG1.0 另一个分支基于生成网络分支的特征计算草图相似度矩阵(affinity matrix), 并通过相似度矩阵与聚类算法^[93]得到组数与分组结果. USPG1.0 基于生成损失函数与由局部和全局聚类损失函数构成的判别损失函数进行训练, 其中, 生成损失函数可用于提高模型的泛化能力, 从而处理新的草图类别, 并使模型可以应用到任意的草图类别上; 判别损失函数则可以保证全局聚类的一致性.

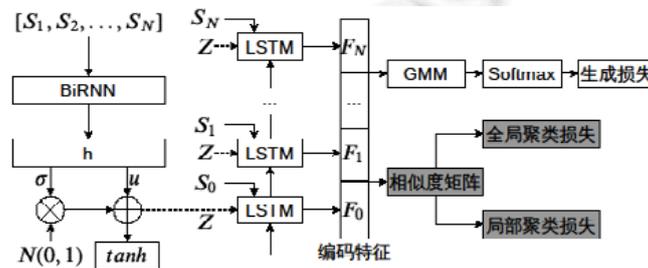


图 15 USPG1.0/USPG2.0 的模型结构图^[34,43]

通过添加额外的分类损失函数, 可以将 USPG2.0 模型改为草图笔画语义分割网络 USGP2.0-SEG-PG; 通

过将模型 USPG2.0 的分组损失函数替换为标签损失函数, 可以得到草图笔画的语义分割模型 USPG2.0-SEG; 另外, 基于草图聚类学习的特征进行精调, 可以得到草图语义分割模型 USPG2.0-SEG+PG+Pre-train. 经过 SPG 数据集验证, USPG2.0-SEG+PG+Pre-train (P -metric 84.6%, C -metric 79.4%) 优于 USGP2.0-SEG-PG (P -metric 83.9%, C -metric 78.8%) 和 USPG2.0-SEG (P -metric 78.5%, C -metric 72.1%). 这表明草图聚类学习到的特征可以提升草图语义分割的性能, 并且具有更好的推广性.

草图聚类算法可以对不同类别的草图进行分组, 基于笔画间的相似性无法保证分割的结果具有语义性. 由于事先无法确定分组的数目与对应关系, 难以直接建立损失函数. 一般通过相似性矩阵对预测结果进行监督, 再基于已有的聚类算法^[93], 在不引入额外参数的情况下获得最终的分组结果. USPG1.0 模型发现: 不引入全局聚类一致性约束时, 草图笔画对之间的相似性存在噪声, 容易导致全局的聚类不一致性. 另外, 草图聚类算法与草图语义分割算法之间的区别与无监督图像分割算法^[94,95]与图像语义分割的区别类似, 通过将相应的感知分组损失替换成分组编号损失或直接添加, 草图聚类算法可转换为草图分割算法^[43].

3.4 像素级标注算法——草图解析

草图解析(sketch parsing)算法的草图分割结果与自然图像的语义分割相似, 是对草图图像的逐像素分类标签. 与草图语义分割相比, 草图解析既分割草图中的笔画, 也分割草图中的空白区域. 理论上, 可以基于自然图像语义分割的算法将输入由自然图像改为草图图像, 并将标注由自然图像语义分割标注改为草图解析标注进行训练, 得到草图解析模型.

Sarvadevabhatla 等人^[49]提出了第一个深度神经网络结构 SketchParse, 对手绘物体草图进行自动解析. 如图 16 所示, SketchParse 模型由两个全卷积网络层构成: 一层为通用的共享网络, 另一层由多个针对特定物体种类的专家网络构成. 两个网络层之间由一个路由网络层(router layer)连接, 它能够利用从共享网络获得的特征选择对应的专家网络, 从而避免依据输入草图的类别手动对 SketchParse 进行调整. 另外, SketchParse 可以与物体姿态估计(object pose prediction)相结合, 将物体姿态估计作为一个附属的学习任务, 从而提高整体的性能. SketchParse 利用自然图像数据集 Pascal-Parts^[96]和 CORE^[51]生成合成草图进行训练, 并在草图数据集^[31,52]上进行了验证. SketchParse 模型能够处理未知但相近的类, 还能提高基于草图的细粒度图像检索(fine-grained sketch-based image retrieval).

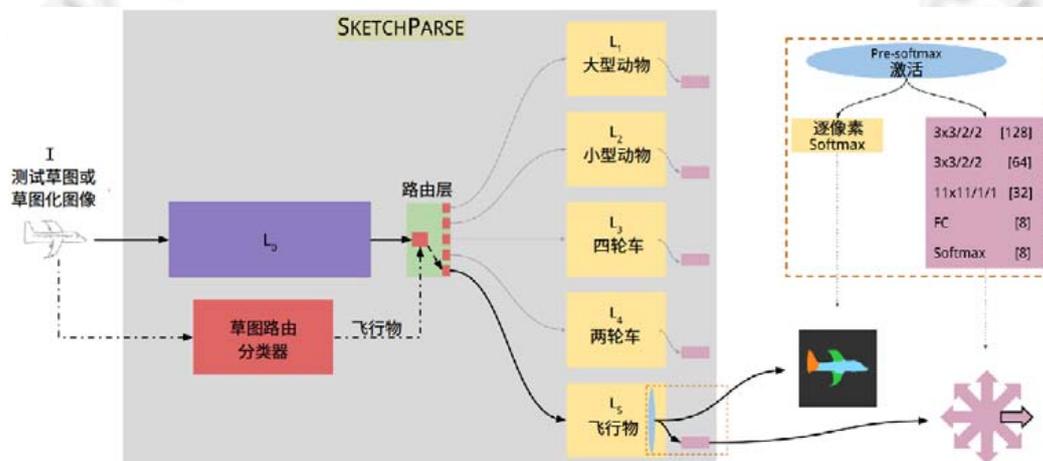


图 16 SketchParse 网络结构^[49]

Zheng 等人^[58]提出了深度神经网络结构 DeepSSP (deep semantic sketch parsing)对手绘草图进行解析: 由于缺乏可直接用于草图解析训练的数据集, DeepSSP 模型利用齐次变换(homogeneous transformation)解决领域迁移(domain adaptation)问题, 从而将自然图像语义数据集应用于训练草图语义解析; 提出了一个软加权损失

函数(soft-weighted loss function)缓解模糊的边界标注与样本不均匀问题;采取分步训练策略(staged learning strategy)充分利用已有的数据集训练共享网络层与专家网络层.

3.5 草图分割方法总结

如图 17 所示,采用相同的语义分割网络结构如 SegNet, DeepLab 等,当训练模型时输入为自然图像,标注为自然图像语义分割的结果时,模型可以用于自然图像的语义分割应用;当训练模型时输入为草图图像,标注为草图图像的笔画级语义分割结果,忽略背景像素,仅利用笔画上的像素进行模型更新,模型可以用草图图像的笔画级语义分割应用;当训练模型时输入为草图图像,标注为草图图像的像素级语义解析结果时,模型可以用于像素级语义解析应用.采用图像作为草图的表达形式,可以充分利用处理自然图像的神经网络模型与数据集.但由于草图特有的稀疏性与多样性,可能会带来前景背景样本不均匀的问题,并使得一些针对上下文信息的网络结构失效^[53].针对草图进行处理的网络结构不同于自然图像,需要侧重于学习线条、边界、轮廓等连接关系特征,而不是纹理和颜色.而前景背景的背景样本不均匀问题一般通过对类别进行加权解决,如通过将背景像素的损失权重设置为 0,即忽略背景像素的损失时,可以极大地提升网络性能^[56].但另一些文献发现:通过简单的比例加权,或者忽略背景像素,对草图分割的结果无明显提升^[55];而通过语义分割网络得到分割结果,再利用图割算法进行优化,弥补当前全卷积语义分割网络的不足,充分利用连接关系,可以有效提高分割的精度.

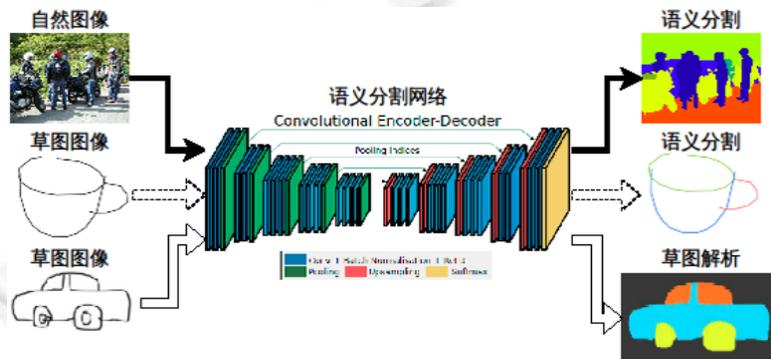


图 17 当通过图像的方式输入草图时,草图语义分割与草图解析可以共享自然图像语义分割网络模型

目前已经提出的草图分割算法主要针对单目标草图语义分割,尽管基于图像的单目标草图语义分割算法可以直接应用于草图场景下的多目标分割,但基于序列与点集的方法需要进行相应的调整,从而应对多目标环境下遮挡与噪声等挑战.草图感知聚类与草图解析算法目前相对较少,其中,草图感知聚类通过学习与利用人类分割物体的原理,如接近性、相似性等,原则上可以在开放环境下对草图笔画进行分组,从而应用到类别未知的场景下,如手工流程图表分割、数据集之外的草图类别分割.而草图解析算法侧重于对区域的分割,适合应用于草图与相关图像编辑的应用,目前已有的方法基于像素,而不是通过草图的笔画对空间进行划分,再获得最终的区域分割结果.

4 实验对比分析

4.1 草图语义分割算法评测

草图笔画语义分割算法主要采用 Huang 等人^[41]提出的评测指标.

- 基于像素的精度 P -metric (pixel-based accuracy), 即像素分类正确的比例.
- 基于部件的精度 C -metric (component-based accuracy), 即组成部件分类正确的比例, 而当一个组成部

件 75%以上的像素都分类正确, 则此部件分类正确.

由于大部件占用像素更多, 因此基于像素的评测方法会倾向于大部件; 而由于小部件通常要多于大部件, 因此基于部件的精度会倾向于小部件^[41]. 文献[44,45]中采用稍微不同的指标.

- 基于笔画的精度(stroke-based accuracy), 即当 80%的像素分割正确时, 笔画分类正确.
- 基于部件的精度(component-based accuracy), 即当 80%的笔画分割正确时, 组件分割正确.

如表 3 所示, 基于深度学习的分割方法普遍优于基于人工设计特征的方法 Huang-Auto^[41]与 Sketch-CRF^[38]. 其中, Sketch-a-Net 采用深度神经网络进行草图识别, 2DSeg^[10]将其输入由整张草图改为单个笔画, 并将其任务由草图识别改为草图语义分割. 由于单个笔画缺乏全局信息, 同时具有模糊性, 即相同笔画在不同草图中具有不同的语义标签, 因此其 *P-metric* 在 Huang 数据集上仅达到 70%, 低于非深度学习分割方法 Sketch-CRF. 2DSeg 与 Sketch-VGG^[57]在输入单个笔画的同时输入整张草图, 使得草图分割的 *P-metric* 在 Huang 数据集上达到 82.8%与 83.4%. 由于 Huang 与 Schneider 数据集的规模较小, 基于图像的方法如 Sketch-VGG 利用迁移学习可以降低对训练数据的数量要求, 获得一定程度上性能提升, 其结果普遍优于其他方法. SPG 与 SketchSeg-150K 数据集规模较大, 基于图卷积与点集的方法如 Sketch-GCN 在需要更多训练数据的同时, 具备更强的特征学习能力, 因此 Sketch-GCN 模型在其上面的 *P-metric* 均超过了 95%. 对于组件部件分割结果的评测 *C-metric* 见表 4, 与表 3 基本一致, 而针对自然图像语义分割而设计的网络模型 DeepLab 在 Huang 数据集上 *P-metric* 上达到中等 76.2%, 但 *C-metric* 仅为 66.7%, 说明 DeepLab 容易在小部件上出现分类错误. USPG2.0 模型则综合感知分组与语义分割, 在 *P-metric* 与 *C-metric* 上均获得了前三的结果. 目前, 草图分割数据集中草图种类有限, 最多的为 SPG 数据集的 25 类, 同时, 草图语义分割需要预先判断草图图像类别再进一步进行细分, 当扩展到更多种类的草图分割时, 实时性与精确度难以保证.

表 3 *P-metric* 指标评测结果, 值越大表明效果越好 (%)

	Huang	Schneider	SPG	SketchSeg-150K
Huang-Auto ^[41]	67.9	-	-	-
Sketch-a-Net ^[9]	70.0	82.3	-	-
Sketch-CRF ^[38]	73.2	83.0	-	-
SFSegNet ^[56]	74.2	-	-	-
MCPNet ^[46]	75.1	-	-	-
Dual-CNN ^[54]	75.7	86.2	-	-
DeepLab ^[14]	62.7-76.2	84.2	80.5	-
Sketch-GCN ^[30]	79.5	89.3	97.6	98
Sketch-GCN+GC ^[30]	81.3	89.8	-	-
FastSeg+GC ^[55]	80.6-81.4	83.6-84.8	86.2	95
2DSeg ^[10]	80.4-82.8	91.8-91.9	-	-
Sketch-VGG ^[57]	82.0-83.4	93.8	-	-
USPG2.0 ^[43]	83.9-84.6	88.8-89.0	84.9	-
SketchSegNet ^[44]	-	-	-	90

表 4 *C-metric* 指标评测结果, 值越大表明效果越好 (%)

	Huang	Schneider	SPG	SketchSeg-150K
Huang-Auto ^[41]	55.3	-	-	-
Sketch-a-Net ^[9]	66.4	77.7	-	-
DeepLab ^[14]	50.0-66.7	78.8	66.1	-
Sketch-CRF ^[38]	67.0	81.8	-	-
Sketch-GCN ^[30]	69.2	82.2	95.6	98
Dual-CNN ^[54]	69.3	78.8	-	-
FastSeg+GC ^[55]	73.8-74.7	78.4-79.4	77.4	92
Sketch-GCN+GC ^[30]	76.0	84.4	-	-
Kaiyrbekov ^[42]	77.1	-	-	-
2DSeg ^[10]	76.1-77.9	90.8-91.2	-	-
USPG2.0 ^[43]	78.8-79.4	85.2-85.8	77.6	-
Sketch-VGG ^[57]	83.3-85.3	92.8	-	-
SketchSegNet ^[44]	-	-	-	88

4.2 草图感知聚类算法评测

草图分组算法主要采用 VOI (variation of information), PRI (probabilistic rand index)和 SC (segmentation covering)^[34,97]指标, 其中, VOI 度量不同分组间的平均条件熵距离, PRI 度量不同分组之间笔画的相容性, SC 度量不同分组之间的重叠度. 如表 5 所示, 在 SPG 与 Edge-PG 数据集上, 基于深度学习的模型 USPG1.0 与 USPG2.0 优于非深度学习方法 Edge-PG. 而模型 DeepLab 采用的语义分割的监督训练方式, 容易在已知类别(训练过的类别)上过拟合, 但 USPG1.0 和 USPG2.0 在已知和未知类别的各个指标上均优于 DeepLab 模型, 同时, 其指标在已知和未知类别上差别较小, USPG2.0 在未知类别上的 VOI 指标甚至优于已知类别, 这说明模型 USPG1.0 和 USPG2.0 对于已知类别具备与语义分割模型类似的特征学习能力, 同时, 对于未知类别具备良好的推广性. USPG 系列模型结合 GAN 训练框架, 对中间特征进行编码, 因而训练可能不稳定, 可以采取更稳定的中间特征编码方式如变分自编码器.

表 5 感知聚类算法评测结果, ↑表示对应指标的值越大则效果越好,
↓表示对应指标的值越小则效果越好

数据集	SPG 已知(未知)类别			Edge-PG		
	VOI↓	PRI↑	SC↑	VOI↓	PRI↑	SC↑
DeepLab ^[14]	1.20(-)	0.73(-)	0.65(-)	-	-	-
Edge-PG ^[40]	1.03(1.13)	0.75(0.69)	0.65(0.61)	1.69	0.62	0.53
USPG1.0 ^[34]	0.59(0.64)	0.87(0.86)	0.79(0.77)	0.96	0.78	0.71
USPG2.0 ^[43]	0.55(0.54)	0.91(0.89)	0.84(0.83)	0.81	0.82	0.76

4.3 草图解析算法评测

草图解析算法主要沿用自然图像语义分割的 IOU (intersection over union)指标. 如表 6 所示, SketchParse 模型基于 DeepLab 与 Sketch-a-Net 模型, 在 SketchParse 数据集上取得了平均 IOU 为 64.45%的良好结果(评测指标取自模型 DeepSSP^[58]的文献,略优于原文结果 63.17%). DeepSSP 则证实, 在草图解析上, 可以利用齐次变换解决迁移学习问题, 并通过软加权损失函数处理样本不均衡问题, 从而提高分割精度. 由于数据集的缺乏, 目前草图解析的工作相对较少. 同时, 草图解析算法与自然图像语义分割方法有很多相似的地方, 主要的区别在于草图解析中的数据由笔画构成, 而前者的精度远低于后者的(自然图像语义分割的平均 IOU 指标达到了 80%以上). 其主要原因为, 草图解析存在更大的样本不均匀问题. 如图 9 所示, 自行车的轮子占据巨大区域, 而自行车架仅由线条构成, 占据区域很小, 因而容易在学习过程中被忽略. 同时, 草图解析中, 一些草图笔画被忽略, 如图 9 中猫的眼睛与胡子, 这些噪声也可能造成草图解析精度的下降.

表 6 草图解析算法评测结果, 值越大表明效果越好 (%)

	Cow	Horse	Cat	Dog	Sheep	Bus	Car	Bicycle	Motorbike	Airplane	Bird	Average
DeepLab ^[14]	66.01	67.77	66.37	67.41	67.37	65.80	63.15	59.15	50.43	52.95	44.57	60.74
SketchParse ^[49]	68.78	69.35	69.60	71.18	70.81	68.00	67.35	62.66	55.04	57.34	50.89	64.45
DeepSSP ^[58]	70.42	72.81	69.94	72.57	72.02	67.88	70.88	63.30	55.69	59.96	54.38	66.25

5 草图分割应用

草图分割为草图应用提供细精度的操作空间, 从而优化应用的性能. Huang 等人^[41]设想利用分割结果, 基于 2D 草图生成 3D 草图, 通过估计 2D 草图中点对应的 3D 位置, 可以结合分割结果进行优化, 最终生成对应的 3D 草图. Harbi 等人^[98]提出, 可以结合草图分割算法对画钟测试结果进行分割, 再结合预定义的规则, 实现对画钟测试结果的自动打分, 最终反映出病人的视觉重建和图形重建能力. Qi 等人^[39]提出通过对自然图像的边缘提取结果进行草图聚类, 移除部分边缘得到“合成草图”, 再将合成草图数据库用于训练基于草图的细精度图像检索, 从而避免对大型草图-自然图像数据集的依赖, 同时保证草图识别和检索等模型的应用性能. 图 18 展示了基于草图聚类的技术将图像转化为草图的过程. Schneider 等人^[38]设想: 在拥有充分标注分割草图的情况下, 可以采用草图分割模型对模型草图进行分割, 得到相应各个部件的草图, 再利用匹配算法得到相

应的 3D 模型部件, 最后利用部件组合(part assembly)算法构建 3D 模型. Kaiyrbekov 等人^[42]设想辅助基于草图分割, 对用户绘制的草图要素进行分割, 再将草图要素对应到自然图像与视频中的要素, 最终实现对图像和视频的检索. 除了以上文献设想的应用, 草图分割一方面还可以通过将草图分割为部件级的组件, 结合 GAN 图像生成与转换技术为草图上色、漫画设计和动画制作等提供部件级的自动编辑功能, 提高草图编辑的效率; 另一方面, 基于草图分割技术, 不仅可以对用户输入的草图进行分割, 还可以对图表中与草图结构相似的组件进行分割, 实现基于草图的图表分割与处理. 如通过绘制湖泊, 再基于草图分割算法从世界地图中分割出所有的湖泊, 进而对湖泊的面积进行统计、对湖泊的颜色进行修改等. 如对于用户绘制的流程图, 可以通过草图分割技术将其中的箭头与区域进行区分, 从而对用户绘制的流程图进行布局与编辑上的优化.

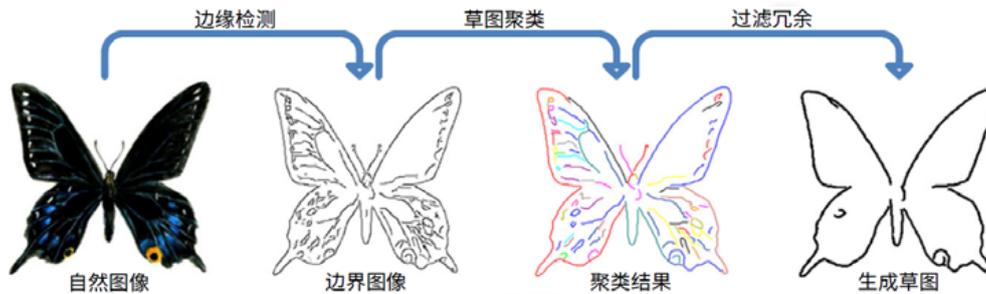


图 18 利用草图聚类由图像生成草图^[39]

6 总 结

本文聚焦草图语义分割、草图聚类与草图解析, 从草图的表示、算法的处理思路、评测方法、数据集和草图分割应用等角度对近年的草图分割研究情况进行了介绍. 基于深度学习的算法在近年来占据了草图分割的主导地位, 基于 GCN, RNN 与点集的方法展现出比 FCN 更好的分割潜力与结果.

草图分割未来可能的方向:

- 1) 充分利用自然图像分割数据集与不带分割标注的草图数据集来训练草图分割模型: 图像与草图具有相似的结构特征, 可以充分应用海量的图像与不带分割标注的草图数据集来避免对大型草图分割标注数据集的依赖.
- 2) 开放环境下的多目标多尺度草图分割: 基于有限类别的单目标草图场景在应用上受到许多限制, 在开放环境下, 可以对未知类别的草图目标进行分割; 在多目标场景下, 需要同时对多种类别的草图目标进行分类与部件级分割.
- 3) 对于复杂草图的分割: 由简单的黑白手绘草图扩展到其他草图, 如彩色的手绘草图、带有区域填充信息的草图或者线条宽度不同的草图、带有不同程度冗余与噪声的草图.
- 4) 特殊硬件上草图分割: 一些特殊硬件上可以获得压力和朝向等额外信息对草图进行分割, 丰富分割的方法, 提高分割的精度与速度.
- 5) 基于手绘草图的分割结合具体应用场景对其他图像进行处理、生成或编辑: 其他图像可以按照用途等分类标准可以粗略地分为统计图表、地图、流程图、水彩画、油画、水墨画、铅笔画、人物画、肖像画等, 其中, 草图分割技术通过预测具有相似语义的结构, 可以将统计图表、地图与流程图分解为具体的模块再进行针对性的处理; 另外, 水彩画、油画等绘画艺术图像与黑白手绘草图与草图存在结构与语义上的联系, 可以针对不同复杂图像的特点与草图简洁的特点相结合, 对不同种类的复杂图像进行部件级的生成与编辑, 如基于分割结果的草图上色与区域填充.

References:

- [1] Song XX, Song DS. The evidence of language iconicity for ancient Chinese character. *Journal of Yunmeng*, 2014, 35(1): 125–128 (in Chinese).
- [2] Ding C, Liu LG. A survey of sketch based modeling systems. *Frontiers of Computer Science*, 2016, 10(6): 985–999.
- [3] Jo Y, Park J. Sc-Fegan: Face editing generative adversarial network with user's sketch and color. In: *Proc. of the IEEE Int'l Conf. on Computer Vision*. 2019. 1745–1753. [doi: 10.1109/ICCV.2019.00183]
- [4] Chen WL, Hays J. Sketchygan: Towards diverse and realistic sketch to image synthesis. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2018. 9416–9425.
- [5] Ghosh A, Zhang R, Dokania P, Wang O, Efros A, Torr P, Shechtman E. Interactive sketch & fill: Multiclass sketch-to-image translation. In: *Proc. of the IEEE Int'l Conf. on Computer Vision*. 2019. 1171–1180. [doi: 10.1109/ICCV.2019.00126]
- [6] Krizhevsky A, Sutskever I, Hinton G. Imagenet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems*. 2012. 1097–1105.
- [7] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv: 1409.1556*, 2014.
- [8] He KM, Zhang XY, Ren SQ, Sun J. Deep residual learning for image recognition. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2016. 770–778.
- [9] Yu Q, Yang YX, Liu F, Song YZ, Xiang T, Hospedales TM. Sketch-a-net: A deep neural network that beats humans. *Int'l Journal of Computer Vision*, 2017, 122(3): 411–425. [doi: 10.1007/s11263-016-0932-3]
- [10] Zhu XY, Xiao Y, Zheng Y. 2D freehand sketch labeling using CNN and CRF. *Multimedia Tools and Applications*, 2020, 79(1): 1585–1602.
- [11] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2015. 3431–3440.
- [12] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: *Proc. of the Int'l Conf. on Medical Image Computing and Computer-assisted Intervention*. Springer, 2015. 234–241.
- [13] Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. Semantic image segmentation with deep convolutional nets and fully connected CRFs. *arXiv: 1412.7062*, 2014.
- [14] Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2017, 40(4): 834–848.
- [15] Chen LC, Papandreou G, Schroff F, Adam H. Rethinking atrous convolution for semantic image segmentation. *arXiv: 1706.05587*, 2017.
- [16] Chen LC, Zhu YK, Papandreou G, Schroff F, Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: *Proc. of the European Conf. on Computer Vision (ECCV)*. 2018. 801–818.
- [17] Jordan MI. Serial order: A parallel distributed processing approach. In: *Proc. of the Advances in Psychology*, Vol.121. Elsevier, 1997. 471–495.
- [18] Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Computation*, 1997, 9(8): 1735–1780.
- [19] Du Y, Wang W, Wang L. Hierarchical recurrent neural network for skeleton based action recognition. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2015. 1110–1118.
- [20] Vinyals O, Toshev A, Bengio S, Erhan D. Show and tell: A neural image caption generator. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2015. 3156–3164.
- [21] Donahue J, Anne Hendricks L, Guadarrama S, Rohrbach M, Venugopalan S, Saenko K, Darrell T. Long-Term recurrent convolutional networks for visual recognition and description. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2015. 2625–2634.
- [22] Sarvadevabhatla RK, Kundu J. Enabling my robot to play pictictionary: Recurrent neural networks for sketch recognition. In: *Proc. of the 24th ACM Int'l Conf. on Multimedia*. 2016. 247–251.
- [23] Ha D, Eck D. A neural representation of sketch drawings. In: *Proc. of the Int'l Conf. on Learning Representations (ICLR 2018)*. 2018. 1–16.

- [24] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Bengio Y. Generative adversarial nets. In: *Advances in Neural Information Processing Systems*. 2014. 2672–2680.
- [25] Wang ZW, She Q, Ward TE. Generative adversarial networks: A survey and taxonomy. arXiv: 1906.01529, 2019.
- [26] Creswell A, Bharath AA. Adversarial training for sketch retrieval. In: *Proc. of the European Conf. on Computer Vision*. Springer, 2016. 798–809.
- [27] Scarselli F, Gori M, Tsoi AC, *et al.* The graph neural network model. *IEEE Trans. on Neural Networks*, 2008, 20(1): 61–80.
- [28] Wu ZH, Pan SR, Chen FW, *et al.* A comprehensive survey on graph neural networks. *IEEE Trans. on Neural Networks and Learning Systems*, 2020, 32(1): 4–24.
- [29] Xu P, Joshi CK, Bresson X. Multi-graph transformer for free-hand sketch recognition. arXiv: 1912.11258, 2019.
- [30] Yang LM, Zhuang JJ, Fu HB, *et al.* SketchGNN: Semantic sketch segmentation with graph neural networks. *ACM Trans. on Graphics (TOG)*, 2021, 40(3): Article No.28.
- [31] Eitz M, Hays J, Alexa M. How do humans sketch objects. *ACM Trans. on Graphics (TOG)*, 2012, 31(4): Article No.44.
- [32] Graves A. Generating sequences with recurrent neural networks. arXiv: 1308.0850, 2013.
- [33] Song JF, Pang KY, Song YZ, *et al.* Learning to sketch with shortcut cycle consistency. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2018. 801–810.
- [34] Li K, Pang KY, Song JF, *et al.* Universal sketch perceptual grouping. In: *Proc. of the European Conf. on Computer Vision (ECCV)*. 2018. 593–609.
- [35] Muhammad UR, Yang YX, Song YZ, *et al.* Learning deep sketch abstraction. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2018. 8014–8023.
- [36] Douglas DH, Peucker TK. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. In: *Proc. of the Classics in Cartography: Reflections on Influential Articles from Cartographica*. 2011. 15–28. [doi: 10.1002/9780470669488.ch2]
- [37] Wang F, Lin SJ, Wu HF, *et al.* Spfusionnet: Sketch segmentation using multi-modal data fusion. In: *Proc. of the 2019 IEEE Int'l Conf. on Multimedia and Expo (ICME)*. 2019. 1654–1659.
- [38] Schneider RG, Tuytelaars T. Example-based sketch segmentation and labeling using CRFs. *ACM Trans. on Graphics (TOG)*, 2016, 35(5): Article No.151.
- [39] Qi YG, Guo J, Song YZ, *et al.* Im2Sketch: Sketch generation by unconflicted perceptual grouping. *Neurocomputing*, 2015, 165: 338–349.
- [40] Qi YG, Song YZ, Xiang T, *et al.* Making better use of edges via perceptual grouping. In: *Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*. 2015. [doi: 10.1109/CVPR.2015.7298795]
- [41] Huang Z, Fu HB, Lau RWH. Data-driven segmentation and labeling of freehand sketches. *ACM Trans. on Graphics (TOG)*, 2014, 33(6): Article No.175.
- [42] Kaiyrbekov K, Sezgin M. Deep stroke-based sketched symbol reconstruction and segmentation. *IEEE Computer Graphics and Applications*, 2020, 40(1): 112–126.
- [43] Li K, Pang KY, Song YZ, *et al.* Toward deep universal sketch perceptual grouper. *IEEE Trans. on Image Processing*, 2019, 28(7): 3219–3231.
- [44] Wu XY, Qi YG, Liu J, *et al.* Sketchsegnet: A RNN model for labeling sketch strokes. In: *Proc. of the 28th IEEE Int'l Workshop on Machine Learning for Signal Processing (MLSP)*. 2018. 1–6.
- [45] Qi YG, Tan ZH. Sketchsegnet+: An end-to-end learning of rnn for multi-class sketch semantic segmentation. *IEEE Access*, 2019, 7: 102717–102726.
- [46] Wang F, Lin SJ, Li HF, *et al.* Multi-column point-CNN for sketch segmentation. *Neurocomputing*, 2020, 392: 50–59.
- [47] Mukherjee K, Hawkins RX, Fan JE. Communicating semantic part information in drawings. In: *Proc. of the Annual Conf. of the Cognitive Science Society*. 2019. 2413–2419.
- [48] Fan JE, Yamins DLK, Turk-Browne NB. Common object representations for visual production and recognition. *Cognitive Science*, 2018, 42(8): 2670–2698.

- [49] Sarvadevabhatla RK, Dwivedi I, Biswas A, *et al.* Sketchparse: Towards rich descriptions for poorly drawn sketches using multi-task hierarchical deep networks. In: Proc. of the 25th ACM Int'l Conf. on Multimedia. 2017. 10–18.
- [50] Chen XJ, Mottaghi R, Liu XB, *et al.* Detect what you can: Detecting and representing objects using holistic models and body parts. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2014. 1971–1978.
- [51] Farhadi A, Endres I, Hoiem D. Attribute-centric recognition for cross-category generalization. In: Proc. of the 2010 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. IEEE, 2010. 2352–2359.
- [52] Sangkloy P, Burnell N, Ham C, *et al.* The sketchy database: Learning to retrieve badly drawn bunnies. ACM Trans. on Graphics (TOG), 2016, 35(4): Article No.119. [doi: 10.1145/2897824.2925954]
- [53] Zou CQ, Yu Q, Du RF, *et al.* Sketchyscene: Richly-annotated scene sketches. In: Proc. of the European Conf. on Computer Vision (ECCV). 2018. 438–454.
- [54] Zhu XY, Xiao Y, Zheng Y. Part-level sketch segmentation and labeling using dual-CNN. In: Proc. of the Int'l Conf. on Neural Information Processing. Springer, 2018. 374–384.
- [55] Li L, Fu HB, Tai CL. Fast sketch segmentation and labeling with deep learning. IEEE Computer Graphics and Applications, 2019, 39(2): 38–51.
- [56] Jiang JK, Wang RM, Lin SJ, *et al.* Sfsegnet: Parse freehand sketches using deep fully convolutional networks. In: Proc. of the 2019 Int'l Joint Conf. on Neural Networks (IJCNN). 2019. 1–8.
- [57] Zhu XY, Yuan J, Xiao Y, *et al.* Stroke classification for sketch segmentation by fine-tuning a developmental VGGNet16. Multimedia Tools and Applications, 2020, 79(45): 33891–33906.
- [58] Zheng Y, Yao HX, Sun XS. Deep semantic parsing of freehand sketches with homogeneous transformation, soft-weighted loss, and staged learning. arXiv: 1910.06023, 2019.
- [59] Xu P. Deep learning for free-hand sketch: A survey. arXiv: 2001.02600, 2020.
- [60] Sezgin TM, Stahovich T, Davis R. Sketch based interfaces: Early processing for sketch understanding. 2001. In: ACM SIGGRAPH 2007 Courses. New York: Association for Computing Machinery, 2007. <https://doi.org/10.1145/1281500.1281548>
- [61] Kim DH, Kim MJ. A curvature estimation for pen input segmentation in sketch-based modeling. Computer-aided Design, 2006, 38(3): 238–248.
- [62] Pu JT, Gur D. Automated freehand sketch segmentation using radial basis functions. Computer Aided Design, 2009, 41(12): 857–864. [doi: 10.1016/j.cad.2009.05.005]
- [63] Sun ZB, Wang CH, Zhang LQ, *et al.* Free hand-drawn sketch segmentation. In: Proc. of the European Conf. on Computer Vision. Berlin, Heidelberg: Springer, 2012. 626–639.
- [64] Noris G, Skora D, Shamir A, *et al.* Smart scribbles for sketch segmentation. In: Proc. of the Computer Graphics Forum, Vol.31. Wiley Online Library, 2012. 2516–2527.
- [65] Lowe DG. Object recognition from local scale-invariant features. In: Proc. of the IEEE Int'l Conf. on Computer Vision. 1999. 1150–1157.
- [66] Sánchez J, Perronnin F, Mensink T, *et al.* Image classification with the fisher vector: Theory and practice. Int'l Journal of Computer Vision, 2013, 105(3): 222–245.
- [67] Lafferty J, McCallum A, Pereira FC. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: Proc. of the Int'l Conf. on Machine Learning (ICML 2001). 2001. 282–289.
- [68] Li SZ. Markov random field models in computer vision. In: Proc. of the European Conf. on Computer Vision. Springer, 1994. 361–370.
- [69] Boykov Y, Veksler O, Zabih R. Fast approximate energy minimization via graph cuts. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2001, 23(11): 1222–1239.
- [70] Jaderberg M, Simonyan K, Zisserman A, *et al.* Spatial transformer networks. In: Proc. of the Advances in Neural Information Processing Systems. 2015. 2017–2025.
- [71] Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481–2495.

- [72] Schuster M, Paliwal KK. Bidirectional recurrent neural networks. *IEEE transactions on Signal Processing*, 1997, 45(11): 2673–2681.
- [73] Graves A, Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks*, 2005, 18(5-6): 602–610.
- [74] Kingma DP, Welling M. Auto-encoding variational bayes. arXiv: 1312.6114, 2013.
- [75] Bowman SR, Vilnis L, Vinyals O, *et al.* Generating sentences from a continuous space. arXiv: 1511.06349, 2015.
- [76] Inselberg A, Dimsdale B. Parallel coordinates for visualizing multi-dimensional geometry. In: *Proc. of the Computer Graphics '87*. Springer, 1987. 25–44.
- [77] Gers FA, Schmidhuber J, Cummins F. Learning to forget: Continual prediction with LSTM. *Neural Computation*, 2000, 12(10): 2451–2471.
- [78] Qi CR, Su H, Mo K, *et al.* Pointnet: Deep learning on point sets for 3d classification and segmentation. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2017. 652–660.
- [79] Wagemans J, Elder JH, Kubovy M, *et al.* A century of gestalt psychology in visual perception: I. perceptual grouping and figure—Ground organization. *Psychological Bulletin*, 2012, 138(6): 1172–1217.
- [80] Elder JH, Goldberg RM. Ecological statistics of gestalt laws for the perceptual organization of contours. *Journal of Vision*, 2002, 2(4): 324–353.
- [81] Wertheimer M. Laws of organization in perceptual forms. In: Ellis WD, ed. *A Source Book of Gestalt Psychology*. Kegan Paul, Trench, Trubner & Company, 1938. 71–88. <https://doi.org/10.1037/11496-005>
- [82] Nan LL, Sharf A, Xie K, *et al.* Conjoining gestalt rules for abstraction of architectural drawings. *ACM Trans. on Graphics (TOG)*, 2011, 30(6): 1–10.
- [83] Qi YG, Guo J, Li Y, *et al.* Sketching by perceptual grouping. In: *Proc. of the 2013 IEEE Int'l Conf. on Image Processing (ICIP 2013)*. 2013. 270–274. [doi: 10.1109/ICIP.2013.6738056]
- [84] Lindlbauer D, Haller M, Hancock M, *et al.* Perceptual grouping: Selection assistance for digital sketching. In: *Proc. of the 2013 ACM Int'l Conf. on Interactive Tabletops and Surfaces (ITS 2013)*. 2013. 51–60. [doi: 10.1145/2512349.2512801]
- [85] Elder JH, Krupnik A, Johnston LA. Contour grouping with prior models. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2003, 25(6): 661–674.
- [86] Adluru N, Latecki LJ, Lakaemper R, *et al.* Contour grouping based on local symmetry. In: *Proc. of the 11th IEEE Int'l Conf. on Computer Vision*. IEEE, 2007. 1–8.
- [87] Stahl JS, Wang S. Edge grouping combining boundary and region information. *IEEE Trans. on Image Processing*, 2007, 16(10): 2590–2606.
- [88] Stahl JS, Wang S. Globally optimal grouping for symmetric closed boundaries by combining boundary and region information. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2008, 30(3): 395–411.
- [89] Papari G, Petkov N. Adaptive pseudo dilation for gestalt edge grouping and contour detection. *IEEE Trans. on Image Processing*, 2008, 17(10): 1950–1962.
- [90] Lun ZL, Zou CQ, Huang HB, *et al.* Learning to group discrete graphical patterns. *ACM Trans. on Graphics (TOG)*, 2017, 36(6): Article No.225.
- [91] Greff K, Rasmus A, Berglund M, *et al.* Tagger: Deep unsupervised perceptual grouping. In: *Proc. of the Advances in Neural Information Processing Systems*. 2016. 4484–4492.
- [92] Kim B, Wang O, Öztireli AC, *et al.* Semantic segmentation for line drawing vectorization using neural networks. In: *Proc. of the Computer Graphics Forum, Vol.37*. Wiley Online Library, 2018. 329–338.
- [93] Yang JW, Parikh D, Batra D. Joint unsupervised learning of deep representations and image clusters. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2016. 5147–5156.
- [94] Xia XD, Kulis B. W-net: A deep model for fully unsupervised image segmentation. arXiv: 1711.08506, 2017.
- [95] Wang CL, Yang B, Liao YW. Unsupervised image segmentation using convolutional autoencoder with total variation regularization as preprocessing. In: *Proc. of the 2017 IEEE Int'l Conf. on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017. 1877–1881.

- [96] Wang P, Shen XH, Lin Z, *et al.* Joint object and part segmentation using deep learned potentials. In: Proc. of the IEEE Int'l Conf. on Computer Vision. 2015. 1573–1581.
- [97] Arbelaez P, Maire M, Fowlkes C, *et al.* Contour detection and hierarchical image segmentation. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2010, 33(5): 898–916.
- [98] Harbi Z, Hicks Y, Setchi R, *et al.* Segmentation of clock drawings based on spatial and temporal features. Procedia Computer Science, 2015, 60: 1640–1648.

附中文参考文献:

- [1] 宋潇潇, 宋德生. 语言象似性的古文字学证据. 云梦学刊, 2014, 35(1): 125–128.



王佳欣(1993—), 男, 博士, 主要研究领域为计算机视觉, 深度视觉, 图像分割.



马翠霞(1975—), 女, 博士, 研究员, 博士生导师, CCF 高级会员, 主要研究领域为人机交互, 媒体大数据可视分析.



朱志亮(1988—), 男, 博士, 讲师, CCF 专业会员, 主要研究领域为图像智能感知与增强, 人机交互.



王宏安(1963—), 男, 博士, 研究员, 博士生导师, 主要研究领域为自然人机交互, 实时智能计算.



邓小明(1980—), 男, 博士, 副研究员, CCF 高级会员, 主要研究领域为计算机视觉, 人机交互.